Reviewer's report

Title: Identification of a candidate prognostic gene signature by transcriptome analysis of matched pre- and post-treatment prostatic biopsies from patients with advanced prostate cancer

Version: 2

Date: 20 October 2014

Reviewer: Roslin Russell

Reviewer's report:

1. Is the question posed by the authors well defined?
   Yes the question is well defined and the paper is very straightforward and concise. They are very clear in stating their limitations. Mostly they discuss feasibility of performing RNA-seq on needle biopsies that are potentially very heterogeneous.

2. Are the methods appropriate and well described?
   All the methods, both wet-lab and bioinformatics, are described in full, including software versions etc. One issue I think I should mention is that gene length bias is an issue in RNA-seq data, in which longer genes have higher counts (at the same expression level) than shorter genes with the same ‘nominal’ expression level. One method that takes this into account is GoSeq and provides the probability that a gene will be differentially expressed based on its length alone. None of their enrichment methods take gene length into account. It may be worth their while to investigate this (Discretionary Revision).

3. Are the data sound?
   Yes, the data are sound. They have followed a good quality control assessment and they also mentioned their attrition rate (two samples were removed due to bad quality). They have submitted their data to GEO. I was unable to find any of the supplementary tables and therefore I am unable to comment on these (Minor Essential Revisions).

4. Do the figures appear to be genuine, i.e. without evidence of manipulation?
   Yes, as far as I'm aware, all the data does appear to be genuine.

5. Does the manuscript adhere to the relevant standards for reporting and data deposition?
   Yes, they have mentioned all the software and versions and they have submitted the data to GEO.

6. Are the discussion and conclusions well balanced and adequately supported by the data?
   Overall the discussion is balanced and well-supported by the data but they could
discuss the sample size issue in more detail. One suggestion could be to use their data to estimate what sample size they require for more meaningful results. The paper reports mainly on how feasible it is to perform RNA-seq analysis on needle biopsies and a third of their samples failed. Could they discuss reasons for this and how the attrition rate could be lowered? Also, there is no mention on whether there is information on treatment response. It would be interesting to compare those that respond to those that do but I do understand that sample size is a limitation here. Also, for the post-treatment biopsies, were there any changes in cellularity when compared to pre-treatment? This may correspond to differences between pre- and post treatment. Could they comment on this (Discretionary Revisions).

7. Are limitations of the work clearly stated?
As I have mentioned already, the limitations are clearly stated.

8. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
Yes these are all clearly stated.

9. Do the title and abstract accurately convey what has been found?
Yes, generally speaking.

10. Is the writing acceptable?
Yes the writing is acceptable, clear and concise.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**
I declare that I have no competing interests.